IJST (2012) A3: 305-310
Iranian Journal of Science & Technology

http://www.shirazu.ac.ir/en

Permutation approach for adjusted Durbin rank test used in balanced incomplete block designs for tied data

E. Y. Gokpinar*, F. Gokpinar and H. Bayrak

Department of Statistics, Faculty of Science, Gazi University, Teknikokullar Ankara, Turkey
E-mail: eyigit@gazi.edu.tr

Abstract

Durbin's rank test is widely used for testing treatment effects in Balanced Incomplete Block Designs (BIBDs) which have wide applications in sensory analysis. This test is failed for BIBDs when ties data occur. An adjusted version of Durbin rank test for this kind of data is given to solve this problem. Chi-square approximation, which is commonly used for this test, is not adequate for small BIBDs. For this case, we investigate permutation approach for adjusted Durbin rank test. Also, in this study the tests used in BIBDs are compared by simulation study for tied data, which have not been discussed in the sensory literature.

Keywords: Balanced incomplete block design; permutation approach; adjusted Durbin rank test

1. Introduction

The judges or consumers compare many products by ranking in sensory analysis. Data assigned by each consumer is well known preference ranking. Considering consumers as blocks and products as treatments we can say that a blocked design has been used. Accordingly the Friedman rank test is the most frequently used nonparametric method for testing treatment effects in randomized complete block design. This test is based on within-block ranking. Sometimes there are too many products for a consumer to reliably compare at one sittings. Due to sensory fatigue inaccurate results may be produced. In this situation BIBDs can be used, so each consumer compares only some of the products. In BIBDs Durbin’s rank test which is based on Friedman-type test is widely used for testing treatment effects. Also, Skillings and Mack [1] introduced a test based on Friedman-type test for randomized (both balanced and unbalanced incomplete) block designs [2]. Hence Durbin’s statistic is also called the Durbin–Skillings–Mack statistic [3]. Bi [4] has discussed some computer intensive methods including Monte Carlo test and permutation test for Durbin rank test. Durbin rank test is failed for BIB data when ties data occur. Best et al. [5] suggest an adjusted procedure for tied data in BIBDs. Chi-square approximation, which is commonly used for this test, is not adequate for small BIBDs.

*Corresponding author
Received: 17 August 2011 / Accepted: 18 October 2011

In this study we investigate permutation approach for adjusted Durbin rank test, especially in small BIBDs, which have not been discussed in the sensory literature. Also, we gave brief definitions of Durbin rank test, adjusted Durbin rank test for tied data and Skilling-Mack tests in Section 2. We compare the performance of these tests with simulation according to type I error rates and powers of tests. A summary of our findings is given in Section 3.

2. Test statistics

In this section, definition of BIBDs and the common terms of test statistics are given. BIBD is the one of the most popular designs used in experimental designs. We have t treatments, b blocks, every block has k treatments and every treatment occurs in r different blocks. Also, every treatment pair occurs exactly λ blocks. (t, k, λ) notation is used for BIBDs. The parameters are not independent. The parameters in a BIBD satisfy the restrictions $\lambda(t-1)=r(k-1)$, $rt=bk$ and $b \geq t$. The BIBDs are called symmetrical BIBD if $t=b$ or equivalently, $r=k$. Otherwise, we called this kind of BIBDs asymmetrical BIBDs.

2.1. Durbin rank test

Durbin rank test is a natural extension for BIBDs of the Friedman statistic applied to the case of a complete randomization block design [6]. In Durbin rank test, the linear effect of ith product, M_i, is given in Eq. (1)

$$M_i = \sqrt{\frac{(t-1)}{rt}}\sum_{j=1}^{k} N_{ij}\, g(j) \tag{1}$$

in which,

$$g(j) = \sqrt{\frac{12}{(k^2-1)}}\left(j - \frac{k+1}{2}\right).$$

The definition of M_i involves a difference between the sample mean rank for product i and its expected value assuming a uniform spread of ranks. The values m_i of M_i separate the products according to mean rankings. The test statistic

$$D = \sum_{i=1}^{t} M_i^2 \tag{2}$$

is the Durbin's rank test that looks for average rank differences between products. The Durbin rank test statistic D has chi-square distribution with t-1 degrees of freedom under null hypothesis [7].

2.2 Adjusted durbin rank test for tied data

The Durbin's rank test analyses the data from BIBD with no ties. This situation forces the judges or consumers to choice ranking situation. When there are ties, the Durbin rank test statistic D no longer has chi-square distribution. In this situation, linear effects, M_i, need adjustment by a factor which depends on g(j) [5]. According to Adjusted Durbin rank test, (AD) statistic can be written as

$$AD = \left(\sum_{i=1}^{t} M_i^2\right)\Big/ a \tag{3}$$

The adjustment factor a is

$$a = g'Ug/rt \tag{4}$$

where $g' = (g(1), g(2), \ldots g(k))$ and the (d,w)th element of $U=(U_{ij})$ counts the times rank d and rank w are tied. If for any judge the ranks $d,\ldots d+m-1$ are tied, then $1/m$ is added to each of the m^2 cells corresponding to the submatrix of U, that is to the elements U_{ij} for i,j=$d,\ldots d+m-1$. Since the matrix is constructed by summing over all judges, it is symmetric. Also, the adjusted effect can be written as $M_i/\sqrt{a}$. AD test has an approximate chi-square distribution with t-1 degrees of freedom under null hypothesis [5].

2.3 Skilling and mack test

Skilling and Mack [1] gave other test statistics in incomplete block designs for tied data. In this paper, we give the rearranged case according to BIBDs. For each block the ranks are given from 1 to k. If ties occur, the average rank is given according to Durbin rank test for tied data. Let r_{ij} be the assigned rank for jth block and ith treatment. If ith treatment is not in the jth block, r_{ij} is taken as $(k+1)/2$. The adjusted treatment effect of ith treatment is:

$$c_i = \sum_{j=1}^{b} \sqrt{\frac{12}{k+1}}\left(r_{ij} - \frac{k+1}{2}\right). \tag{5}$$

The variances of c_i (i=1,2,…,t) and the covariances between c_i and c_j (i≠j) can written as follows:

$$Var(c_i) = (t-1)\lambda \quad \text{for i=1,2,}\ldots t. \tag{6}$$

$$Cov(c_i, c_j) = -\lambda \quad \text{for i}\neq\text{j=1,2,}\ldots t \tag{7}$$

Let **c** be the column vector of treatments effects c_i (i=1, 2,…t). The variance-covariance matrix of **c** is defined as:

$$G = \begin{bmatrix} Var(c_1) & Cov(c_1,c_2) & \cdots & Cov(c_1,c_t) \\ Cov(c_2,c_1) & Var(c_2) & \cdots & Cov(c_2,c_t) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(c_t,c_1) & Cov(c_t,c_2) & \cdots & Var(c_t) \end{bmatrix} = \begin{bmatrix} (t-1)\lambda & -\lambda & \cdots & -\lambda \\ -\lambda & (t-1)\lambda & \cdots & -\lambda \\ \vdots & \vdots & \ddots & \vdots \\ -\lambda & -\lambda & \cdots & (t-1)\lambda \end{bmatrix} \tag{8}$$

The test statistic by using Eq. (5) and Eq. (8) is written as $T = \mathbf{c}'G^{-}\mathbf{c}$, where G^{-} is the generalized inverse of G. It can be seen that the rank of G is t-1. So T statistic has approximately chi-square distribution with t-1 degrees of freedom under null hypothesis.

2.4. Permutation approach for adjusted durbin rank test

Fisher [8] firstly proposed permutation approach for experimental design. Bi [4] has discussed this approach for Durbin rank statistic on ranks and BIBDs in the sensory literature. Permutation approach can be used when we do not know the sampling distribution of a test statistic. To estimate the sampling distribution of the test statistic we need many samples generated under the null hypothesis. If the null hypothesis is true, changing the exposure would have no effect on the outcome. By randomly shuffling the exposures we can make up as many data sets as we like. The ranking of the real test statistic among the shuffled test statistics gives a p-value.

Best et al. [5] gave an adjusted version of Durbin rank test for tied data. But chi-square approximation, which is commonly used for this test, is not adequate for small BIBDs, as in this case, we investigate permutation approach for adjusted Durbin rank test (*ADP*). The algorithm for calculating p-value using the Permutation approach could be given as shown below. Algorithm:

1) Compute *AD* test statistic in Eq. (3) for the original data.
2) Choose permutation resample from the data without replacement in a way that is consistent with the null hypothesis of the test and with the study design. By the same way, generate artificial sample a large number of times (say, *M* times)
3) For each of these replicated samples, recalculated the *AD* test statistic in Eq. (3).
4) Let these recalculated *AD* test statistic values be $AD_1^*,...,AD_M^*$. So the permutation distribution of the test statistic is found.
5) Calculate the p value as: p-value= $\frac{\#\left(AD_i^* > AD\right)}{M}$, i=1,…,*M*. Reject the null hypothesis of no treatment effects if p<α and accept null hypothesis otherwise.

3. Simulation study

In this section, we compared the *D, AD, T* and *ADP* tests according to their simulated type 1 error rates and powers of these tests using the Matlab code. For this simulation study symmetric with 2-(4u-1,2u-1,u-1) (u=2,3,4,5,6) parameters and asymmetric with 2-(2u, u, u-1) (u=3,4,5,6) parameters BIBDs are considered. This selected design, which has the highest efficiency in BIBDs is widely used in practice [9].

The designs are generated from using Hadamard matrices. The data for BIBD is generated from binomial with n=9 and p_i (i=1, 2,…t) for each treatment. So this corresponds to the judgment which gives the product rank 1 to 9. For nominal level α=0.05 we used 5000 runs for design to calculate the simulated type 1 error rates and powers of these tests. Table 1-2 display the simulated type 1 error rates of tests for selected designs.

From Table 1, one can see that, the *D* and *T* tests seem to have a type I error rate lower than the nominal level. The type I error rate of *AD* test is close to the nominal level as moderately large BIBDs. The *ADP* test performs superior to the other tests for especially small parameters of designs.

Table 1. Estimated type 1 error rates of tests for symmetric designs for α=0.05

Design	Tests			
2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
2-(7,3,1)	0.000	0.002	0.000	0.044
2-(11,5,2)	0.005	0.035	0.018	0.052
2-(15,7,3)	0.004	0.036	0.019	0.050
2-(19,9,4)	0.012	0.050	0.027	0.050
2-(23,11,5)	0.013	0.050	0.028	0.050

Table 2. Estimated type 1 error rates of tests for asymmetric designs for α=0.05

Design	Tests			
2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
2-(6,3,2)	0.007	0.032	0.013	0.063
2-(8,4,3)	0.011	0.039	0.024	0.065
2-(10,5,4)	0.009	0.044	0.021	0.056
2-(12,6,5)	0.009	0.047	0.022	0.054

From Table 2, it can be see that, the *D* and *T* tests have similar results as Table 1. While the *AD* test seems to have a type I error rate lower than the nominal level, the *ADP* test exceeds the nominal level for small BIBDs. But the difference between type I error and nominal level of these tests is close to each other.

The *ADP* test generally performs better than other tests, especially small BIBDs for symmetric and asymmetric design.

We also compare the *D, AD, T* and *ADP* tests when products are divided into two and three different groups with different binomial probabilities. We take differences between binomial probabilities of group 0.05, 0.10 and 0.20. So the rejection rate of each testing procedure is calculated and compared with the nominal level when the treatment effects are not all equal. When treatments were divided into two groups, Tables 3-4 display the power of tests for symmetric and asymmetric design, respectively. When treatments were divided into three groups Table 5-6 display the power of tests for symmetric and asymmetric design, respectively.

Table 3. Power of tests for symmetric design when treatments were divided into two groups

Binomial probabilities	Design 2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
0.15, 0.25	2-(7,3,1)	0.0005	0.0095	0.0005	0.0865
	2-(11,5,2)	0.0360	0.1706	0.1046	0.2361
	2-(15,7,3)	0.1381	0.4037	0.2861	0.4487
	2-(19,9,4)	0.3240	0.6660	0.5430	0.6970
	2-(23,11,5)	0.5788	0.8389	0.7734	0.8524
0.15, 0.35	2-(7,3,1)	0.0005	0.0265	0.0015	0.2386
	2-(11,5,2)	0.3042	0.6143	0.5503	0.7134
	2-(15,7,3)	0.8779	0.9710	0.9640	0.9775
	2-(19,9,4)	0.9940	0.9995	0.9995	0.9995
	2-(23,11,5)	1.0000	1.0000	1.0000	1.0000
0.45, 0.55	2-(7,3,1)	0.0005	0.0060	0.0010	0.0775
	2-(11,5,2)	0.0250	0.1141	0.0735	0.1621
	2-(15,7,3)	0.0775	0.2436	0.1801	0.2871
	2-(19,9,4)	0.1770	0.4070	0.3440	0.4410
	2-(23,11,5)	0.3527	0.5958	0.5448	0.6268
0.40, 0.60	2-(7,3,1)	0.0005	0.0185	0.0010	0.1821
	2-(11,5,2)	0.2256	0.4742	0.4082	0.5698
	2-(15,7,3)	0.6923	0.8859	0.8649	0.9065
	2-(19,9,4)	0.9590	0.9910	0.9900	0.9930
	2-(23,11,5)	0.9980	1.0000	1.0000	1.0000

Table 4. Power of tests for asymmetric design when treatments were divided into two groups

Binomial probabilities	Design 2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
0.15, 0.25	2-(6,3,2)	0.1033	0.1795	0.0505	0.1770
	2-(8,4,3)	0.1963	0.2793	0.0815	0.2760
	2-(10,5,4)	0.3948	0.4388	0.1320	0.4384
	2-(12,6,5)	0.5820	0.6302	0.2596	0.6302
0.15, 0.35	2-(6,3,2)	0.3668	0.5144	0.1718	0.5092
	2-(8,4,3)	0.7798	0.8450	0.5232	0.8452
	2-(10,5,4)	0.9604	0.9702	0.8278	0.9706
	2-(12,6,5)	0.9960	0.9970	0.9725	0.9970
0.45, 0.55	2-(6,3,2)	0.0666	0.1252	0.0220	0.1220
	2-(8,4,3)	0.1418	0.2102	0.0512	0.2100
	2-(10,5,4)	0.2505	0.2955	0.0845	0.2920
	2-(12,6,5)	0.3800	0.4340	0.1600	0.4400
0.40, 0.60	2-(6,3,2)	0.2610	0.3858	0.1322	0.3806
	2-(8,4,3)	0.6166	0.7072	0.3948	0.7078
	2-(10,5,4)	0.8730	0.8970	0.6836	0.8968
	2-(12,6,5)	0.9768	0.9840	0.9024	0.9840

Table 5. Power of tests for symmetric design when treatments were divided into three groups

Binomial probabilities	Design 2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
0.15, 0.20, 0.25	2-(7,3,1)	0.000	0.007	0.002	0.079
	2-(11,5,2)	0.019	0.105	0.062	0.157
	2-(15,7,3)	0.058	0.248	0.158	0.292
	2-(19,9,4)	0.161	0.439	0.341	0.473
	2-(23,11,5)	0.316	0.638	0.529	0.663
0.15, 0.25, 0.35	2-(7,3,1)	0.000	0.013	0.002	0.173
	2-(11,5,2)	0.165	0.438	0.347	0.523
	2-(15,7,3)	0.576	0.830	0.804	0.860
	2-(19,9,4)	0.911	0.984	0.979	0.987
	2-(23,11,5)	0.996	0.999	0.999	0.999
0.45, 0.50, 0.55	2-(7,3,1)	0.000	0.004	0.000	0.073
	2-(11,5,2)	0.021	0.073	0.048	0.117
	2-(15,7,3)	0.044	0.162	0.119	0.200
	2-(19,9,4)	0.076	0.246	0.193	0.279
	2-(23,11,5)	0.167	0.403	0.332	0.425
0.40, 0.50, 0.60	2-(7,3,1)	0.000	0.007	0.001	0.147
	2-(11,5,2)	0.112	0.302	0.243	0.393
	2-(15,7,3)	0.387	0.659	0.629	0.705
	2-(19,9,4)	0.780	0.931	0.920	0.940
	2-(23,11,5)	0.959	0.993	0.994	0.994

Table 6. Power of tests for asymmetric design when treatments were divided into three groups

Binomial probabilities	Design 2-(4u-1,2u-1,u-1)	*D*	*AD*	*T*	*ADP*
0.15, 0.20, 0.25	2-(6,3,2)	0.015	0.071	0.029	0.124
	2-(8,4,3)	0.031	0.146	0.087	0.215
	2-(10,5,4)	0.073	0.258	0.177	0.301
	2-(12,6,5)	0.130	0.387	0.298	0.443
0.15, 0.25, 0.35	2-(6,3,2)	0.081	0.222	0.123	0.347
	2-(8,4,3)	0.295	0.568	0.480	0.669
	2-(10,5,4)	0.593	0.853	0.806	0.884
	2-(12,6,5)	0.826	0.957	0.956	0.970
0.45, 0.50, 0.55	2-(6,3,2)	0.016	0.061	0.025	0.109
	2-(8,4,3)	0.029	0.102	0.059	0.148
	2-(10,5,4)	0.053	0.148	0.123	0.187
	2-(12,6,5)	0.076	0.249	0.194	0.297
0.40, 0.50, 0.60	2-(6,3,2)	0.060	0.141	0.082	0.243
	2-(8,4,3)	0.206	0.410	0.332	0.519
	2-(10,5,4)	0.442	0.687	0.636	0.724
	2-(12,6,5)	0.661	0.867	0.847	0.887

We observe the following from the numerical results in Table 3-6. The powers of *A, AD, T* tests are not high enough, especially in small BIBDs. The *ADP* test is superior to other tests for symmetric and asymmetric design, especially small BIBDs. Also, the *AD* test appears to be more powerful than the *D* and *T* tests in this situation. While *ADP* test is more powerful than *AD* test in Table 3 the *ADP* and *AD* tests exhibit close power properties, especially small BIBDs in Table 4. When the block size increases, i.e. large BIBDs, it is seen that the power of all three tests is getting higher. In large BIBDs, the *AD* and *ADP* tests exhibit close power properties that appear to be more powerful than the other tests. However, the difference between the powers of the *AD* and *ADP* tests for symmetric design are much higher than the difference between the powers of these tests for asymmetric design. For example, while the difference between these tests is 0.072 for (7, 3, 1) design from Table 5, this difference is 0.053 for (6, 3, 2). It shows that the power of the *ADP* tests for symmetric design is much higher than its power for asymmetric design.

As expected, the powers of all these tests increase when the differences between binomial probabilities of groups are increased. In the same differences between binomial probabilities, the powers of all these tests increase when these probabilities are at extremes in most situations. In all conditions it is seen that *ADP* test is superior to other tests.

4. Conclusion

We studied some nonparametric tests for ranked data with ties data under balanced incomplete block design. Also, distribution of *AD* test was obtained via permutation approach. We compared the performance of the defined tests according to the size and power of the tests.

Finally, it can be said that the *ADP* test appears to be more powerful than the other tests, especially small parameters of designs. Also, in this study the tests used in BIBDs are compared by simulation study for tied data, which has not been discussed in the sensory literature. Although the size and power comparisons are limited they illustrate that the *ADP* test is preferable to *AD* test for tied data in small BIBDs.

References

[1] Skillings, J. H. & Mack, G. A. (1981). On the use of a Friedman type statistic in balanced and unbalanced block designs. *Technometrics*, *23*, 171-177.

[2] Gisbrecht, F. G. & Gumbertz, M. L. (2004). *Planning, Construction and Statistical Analysis of Comparitive Experimemts*. New York, Wiley.

[3] Hollander, M. & Wolfe, D. A. (1999). *Nonparametric statistical methods*. 2nd ed., New York, Wiley.

[4] Bi, J. (2009). Computer-intensive methods for sensory data analysis, exemplified by Durbin's rank test. *Food Quality and Preference*, *20*(3), 195-202.

[5] Best, D. J., Brockhoff, P. B. & Rayner J. C. W. (2006). Tests for balanced incomplete block ranked data with ties, *60*(1), 3-11.

[6] Avlo, M. & Cabilo, P. (1998). Application of Hamming distance to the analysis of block designs. Asymptotic methods in probability and statistics, *Elsevier Science*, Amsterdam, 787-800.

[7] Durbin, J. (1951). Incomplete blocks in ranking experiments. *British Journal of Psychology (Statistical Section)* 4, 85-90.

[8] Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd, Edinburgh.

[9] Raghavarao, D. (1971). Constructions and Combinatorial Problems in Design of Experiments. New York, Wiley.